

SEQUENCE LISTING

5 (1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Hoechst Aktiengesellschaft
 (B) STREET: -
 10 (C) CITY: Frankfurt
 (D) STATE: -
 (E) COUNTRY: Germany
 (F) POSTAL CODE (ZIP): 65926
 (G) TELEPHONE: 069-305-7072
 15 (H) TELEFAX: 069-35-7175
 (I) TELEX: -

(ii) TITLE OF INVENTION: Purification of higher order transcription
 complexes from transgenic non-human animals

20 (iii) NUMBER OF SEQUENCES: 17

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
 25 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

30 (2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 35 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

40 (ix) FEATURE:

(A) NAME/KEY: Peptide
 (B) LOCATION: 1..12

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Val
 1 5 10

50 (2) INFORMATION FOR SEQ ID NO: 2:

55 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids

(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 1..22

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGAGCAACCG CCTGCTGGGT GC

22

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 1..21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCTGTGTTGC CTGCTGGGAC G

21

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGAGACTGAA GTTAGGCCAG C

21

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..76

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCGGCACCAG GCCGCTGCTG TGATGATGAT GATGATGGCT GCTGCCCATG ACTGCGTAAT
GCGGTCATGA CGCTTT

60

76

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..75

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAAGGGGGTG GGGGAGGCAA GGGTACATGA GAGCCATTAC GTCGTCTTCC TGAATCCCTT

60

TAGCCGCTTT GCTCG

75

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..22

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCCTATGACG TCCCGGATTA CG

22

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..22

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTGGAGTGGT GCCCGCAAG GG

22

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
 (A) NAME/KEY: Peptide
 (B) LOCATION: 1..19

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
 1 5 10 15
 Arg Gly Cys

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1310 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
 (A) NAME/KEY: exon
 (B) LOCATION: 1..1310

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCATGGGCTA TCCCTATGAC GTCCCGGATT ACGCAGTCAT GGGCAGCAGC CATCATCATC 60
 ATCATCACAG CAGCGGCCTG GTGCCGCGCG GCAGCCATAT GGATCAGAAC AACAGCCTGC 120
 CACCTTACGC TCAGGGCTTG GCCTCCCCCTC AGGGTGCCAT GACTCCCGGA ATCCCTATCT 180
 TTAGTCCAAT GATGCCTTAT GGCCTGGAC TGACCCCAACA GCCTATTTCAG AACACCAATA 240
 GTCTGTCTAT TTTGGAAGAG CAACAAAGGC AGCAGCAGCA ACAACAACAG CAGCAGCAGC 300
 AGCAGCAGCA GCAGCAACAG CAACAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC 360
 AGCAGCAGCA GCAGCAGCAA CAGGCAGTGG CAGCTGCAGC CGTTCAGCAG TCAACGTCCC 420
 AGCAGGCAAC ACAGGGAACC TCAGGCCAGG CACCACAGCT CTTCCACTCA CAGACTCTCA 480
 CAACTGCACC CTTGCCGGGC ACCACTCCAC TGTATCCCTC CCCATGACT CCCATGACCC 540
 CCATCACTCC TGCCACGCCA GCTTCGGAGA GTTCTGGGAT TGTACCGCAG CTGCAAAATA 600
 TTGTATCCAC AGTGAATCTT GGTGTGTAAC TTGACCTAAA GACCATTGCA CTTCGTGCCC 660
 GAAACGCCGA ATATAATCCC AAGCGGTTTG CTGCGGTAAT CATGAGGATA AGAGAGCCAC 720

GAACCACGGC ACTGATTTTC AGTTCTGGGA AAATGGTGTG CACAGGAGCC AAGAGTGAAG 780
AACAGTCCAG ACTGGCAGCA AGAAAATATG CTAGAGTTGT ACAGAAGTTG GGTTTTCCAG 840
5 CTAAGTTCTT GGAAGTTCAAG ATTCAGAACA TGGTGGGGAG CTGTGATGTG AAGTTTCCTA 900
TAAGGTTAGA AGGCCTTGTG CTCACCCACC AACAATTTAG TAGTTATGAG CCAGAGTTAT 960
10 TTCCTGGTTT AATCTACAGA ATGATCAAAC CCAGAATTGT TCTCCTTATT TTTGTTTCTG 1020
GAAAAGTTGT ATTAACAGGT GCTAAAGTCA GAGCAGAAAT TTATGAAGCA TTTGAAAACA 1080
TCTACCCTAT TCTAAAGGGA TTCAGGAAGA CGACGTAATG GCTCTCATGT ACCCTTGCCT 1140
15 CCCCCACCCC CTTCTTTTTT TTTTTTTAAA CAAATCAGTT TGTTTTGGTA CCTTTAAATG 1200
GTGGTGTGTG GAGAAGATGG ATGTTGAGTT GCAGGGTGTG GCACCAGGTG ATGCCCTTCT 1260
GTAAGTGCCC CTTCCGGCAT CCCGGAATTC CTGCAGCCCA ACGCGGCCGC 1310

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4286 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..4286

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GAATTCCCCT GCAGGTCAGT TAGCGTTGGC CACATAGTAG GTTCTCAAAT ACTTGTTAAT 60
AAATAAGTTT GTTCGAGAAG CTGGGCAATG ATATTCTACA GCTGGAAGAA GAAACATAAT 120
GATCTAGTAA TTAGCTCAAT TAAAAATAAA CGTTCCTCTT TCCTCAGAGG AGCATTTCCC 180
45 AAGGCCTGCC TTGATAGCCA TCCAAAAGG CCAAGCTCAT CCAATCTTGC CCTAGATTTA 240
TGCTAAAATG CAGTTACAAT CGATAGGATG ACAGAAAACG ACAGCACTTA TTTAAATATA 300
50 ATAGGCACTT ATTTAAATAG GAGAAGCTGT GACTTCATAG CAAGTGTTGG GGTTAGGAAA 360
CTGGGTGGAT AACTTGCTG ATGCTGTAGA TCTTAGCCTC TACATGAGAT CATGTGGAAA 420
ATCTGAAAGC ATTTTAGGTT CCTTATGTTT GCAATCAAAT AACTGTACAC CTTTAAATTT 480
55 AAAAAGTACC ATGAGGCACA CACACACACT CGCAGGAAC TTTTGGCGTA ACAAACACTAG 540

T02050-426486

	AATTAGATCT AAAAGCTAAC TGTAGGACTG AGTCTATTCT AAAGTCAAAG CCTGGACATC	600
5	TGGAGTACCA GGGGGAGATG ACGTGTTACG GGCTTCCATA AAAGCAGCTG GCTTTGAATG	660
	GAAGGAGCCA AGAGGCCAGC ACAGGAGCGG ATTCGTCGCT TTCACGGCCA TCGAGCCGAA	720
	CCTCTCGCAA GTCCGTGAGC CGTTAAGGAG GCGCCAGTC CCGACCCCTC GCGCCAAGCC	780
10	CCTCGGGGTC CCCGGGCCCTG GTACTCCTTG CCACACGGGA GGGGCGCGGA AGCCGGGGCG	840
	GAGGAGGAGC CAACCCCGGG CTGGGCTGAG ACCCGCAGAG GAAGACGCTC TAGGGATTTG	900
15	TCCCGGACTA GCGAGATGGC AAGGCTGAGG ACGGGAGGCT GATTGAGAGG CGAAGGTACA	960
	CCCTAATCTC AATACAACCT TTGGAGCTAA GCCAGCAATG GTAGAGGGAA GATTCTGCAC	1020
	GTCCCTTCCA GCGGGCCTCC CCGTCACCAC CCGCCCAAC CCGCCCGAC CGGAGCTGAG	1080
20	AGTAATTCAT ACAAAGGAC TCGCCCTGCT CTTGGGGAAT CCCAGGGACC GTCGTTAAAC	1140
	TCCCACTAAC GTAGAACCCA GAGATCGCTG CGTTCCCGCC CCCTCACCCG CCCGCTCTCG	1200
25	TCATCACTGA GGTGGAGAAG AGCATGCGTG AGGCTCCGGT GCGCGTCAGT GGGCAGAGCG	1260
	CACATCGCCC ACAGTCCCCG AGAAGTTGGG GGGAGGGGTC GGCAATTGAA CCGGTGCCTA	1320
	GAGAAGGTGG CGCGGGGTAA ACTGGGAAAG TGATGTCGTG TACTGGCTCC GCCTTTTTTCC	1380
30	CGAGGGTGGG GGAGAACCGT ATATAAGTGC AGTAGTCGCC GTGAACGTTT TTTTTCGCAA	1440
	CGGGTTTGCC GCCAGAACAC AGGTAAAGTGC CGTGTGTGGT TCCCGCGGGC CTGGCCTCTT	1500
35	TACGGGTTAT GGCCCTTGCG TGCCCTGAAT TACTTCCACG CCCCTGGCTG CAGTACGTGA	1560
	TTCTTGATCC CGAGCTTCGG GTTGGAAAGT GGTGGGAGAG TTCGAGGCCT TGCGCTTAAG	1620
	GAGCCCTTC GCCTCGTGCT TGAGTTGAGG CCTGGCCTGG GCGCTGGGGC CGCCGCGTGC	1680
40	GAATCTGGTG GCACCTTCGC GCCTGTCTCG CTGCTTTCGA TAAGTCTCTA GCCATTTAAA	1740
	ATTTTGTGATG ACCTGCTGCG ACGCTTTTTT TCTGGCAAGA TAGTCTTGTA AATGCGGGCC	1800
45	AAGATCTGCA CACTGGTATT TCGGTTTTTG GGGCCGCGGG CGGCGACGGG GCGCGTGCCT	1860
	CCCAGCGCAC ATGTTCCGGC AGGCGGGGCC TGCGAGCGCG GCCACCGAGA ATCGGACGGG	1920
	GGTAGTCTCA AGCTGGCCGG CCTGCTCTGG TGCCTGGCCT CGCGCCGCCG TGTATCGCCC	1980
50	CGCCCTGGGC GGCAAGGCTG GCGCGGTCGG CACCAGTTGC GTGAGCGGAA AGATGGCCGC	2040
	TTCCCGGCCC TGCTGCAGGG AGCTCAAAAT GGAGGACGCG GCGCTCGGGA GAGCGGGCGG	2100
55	GTGAGTCACC CACACAAAGG AAAAGGGCCT TTCCGTCCTC AGCCGTCGCT TCATGTGACT	2160
	CCACGGAGTA CCGGGCGCCG TCCAGGCACC TCGATTAGTT CTCGAGCTTT TGGAGTACGT	2220

CGTCTTTAGG TTGGGGGGAG GGGTTTTATG CGATGGAGTT TCCCCACACT GAGTGGGTGG 2280

AGACTGAAGT TAGGCCAGCT TGGCACTTGA TGTAATTCTC CTTGGAATTT GCCCTTTTTG 2340

5 AGTTTGGATC TTGGTTCATT CTCAAGCCTC AGACAGTGGT TCAAAGTTTT TTTCTTCCAT 2400

TTCAGGTGTC GTGAGGAATT GCCCCGGGGA TCCATGGGCT ATCCCTATGA CGTCCCGGAT 2460

10 TACGCAGTCA TGGGCAGCAG CCATCATCAT CATCATCACA GCAGCGGCCT GGTGCCGCGC 2520

GGCAGCCATA TGGATCAGAA CAACAGCCTG CCACCTTACG CTCAGGGCTT GGCCTCCCCT 2580

15 CAGGGTGCCA TGACTCCCGG AATCCCTATC TTTAGTCCAA TGATGCCTTA TGGCACTGGA 2640

CTGACCCAC AGCCTATTCA GAACACCAAT AGTCTGTCTA TTTTGGAAGA GCAACAAAGG 2700

CAGCAGCAGC AACAACAACA GCAGCAGCAG CAGCAGCAGC AGCAGCAACA GCAACAGCAG 2760

20 CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGGCAGTG 2820

GCAGCTGCAG CCGTTCAGCA GTCAACGTCC CAGCAGGCAA CACAGGGAAC CTCAGGCCAG 2880

25 GCACCACAGC TCCTTCCACTC ACAGACTCTC ACAACTGCAC CCTTGCCGGG CACCACTCCA 2940

CTGTATCCCT CCCCATGAC TCCCATGACC CCCATCACTC CTGCCACGCC AGCTTCGGAG 3000

AGTTCTGGGA TTGTACCGCA GCTGCAAAAT ATTGTATCCA CAGTGAATCT TGGTTGTAAA 3060

30 CTTGACCTAA AGACCAATTGC ACTTCGTGCC CGAAACGCCG AATATAATCC CAAGCGGTTT 3120

GCTGCGGTAA TCATGAGGAT AAGAGAGCCA CGAACCACGG CACTGATTTT CAGTTCTGGG 3180

35 AAAATGGTGT GCACAGGAGC CAAGAGTGAA GAACAGTCCA GACTGGCAGC AAGAAAATAT 3240

GCTAGAGTTG TACAGAAGTT GGGTTTTCCA GCTAAGTTCT TGGACTTCAA GATTCAGAAC 3300

ATGGTGGGGA GCTGTGATGT GAAGTTTCCT ATAAGGTTAG AAGGCCTTGT GCTCACCAC 3360

40 CAACAATTTA GTAGTTATGA GCCAGAGTTA TTTCTTGGTT TAATCTACAG AATGATCAAA 3420

CCCAGAAFTG TTCTCCTTAT TTTTGTCTTCT GGAAAAGTTG TATTAACAGG TGCTAAAGTC 3480

45 AGAGCAGAAA TTTATGAAGC ATTTGAAAAC ATCTACCCTA TTCTAAAGGG ATTCAGGAAG 3540

ACGACGTAAT GGCTCTCATG TACCCTTGCC TCCCCACCC CTTTCTTTTT TTTTTTTTAA 3600

ACAAATCAGT TTGTTTTGGT ACCTTTAAAT GGTGGTGTG TGAGAAGATG GATGTTGAGT 3660

50 TGCAGGGTGT GGCACCAGGT GATGCCCTTC TGTAAGTGCC CCTTCCGGCA TCCCGGATAT 3720

CCTGCAGCCC AACACGGCCG CTCGAGCATG CATCTAGAGA ACGTCACGGC CGCGATCCCC 3780

55 CTGTGCCCTC TAGTTGCCAG CCATCTGGTT GTTTGCCCCT CCCCCGTGCC TTCCTTGACC 3840

CTGGAAGGTG CCACTCCCAC TGTCTTTTCC TAATAAAATG AGGAAATTGC ATCGCAITGT 3900

CTGAGTAGGT GTCATTCTAT TCTGGGGGGT GGGGTGGGGC AGGACAGCAA GGGGGAGGAT 3960
TGGGAAGACA ATAGCAGGCA TGCTGGGGAT GCGGTGGGCT CTATGGGTAC CCAGGTGCTG 4020
5 AAGAATTGAC CCGGTTCTTC CTGGGCCAGA AAGAAGCAGG CACATCCCCT TCTCTGTGAC 4080
ACACCCTGTC CACGCCCCCTG GTTCTTAGTT CCAGCCCCAC TCATAGGACA CTCAACTTGG 4140
10 AGCGGTCTCT CCCTCCCTCA TCAGCCCACC AAACCAAACC TAGCCTCCAA GAGTGGGAAG 4200
AAATTAAAGC AAGAAGGCTA TTAAGTGCAG AGGGAGAGAA AATGCCTCCA ACATGTGAGG 4260
AAGTAATGAT AGAAATCATA GAATTC 4286
15

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3263 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..3263

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ATCGATAAGC TGAGATCCGG CTAGAACTG CTGAGGGCTG GACCGCATCT GGGGACCATC 60
TGTTCTTGGC CCTGAGCGGG GCAGGAACTG CTTACCGCAG ATATCCTGTT TGCCCCAATT 120
CAGCTGTTCC ATCTGTTCTT GGCCCTGAGC GGGGCAGGAA CTGCTTACCA CAGATATCCT 180
40 GTTTGGCCCA TATTCAGCTG TCTCTCTGTT CCTGACCTTG ATCTGAACTT CTCTATTCTC 240
AGTTATGTAT TTTTCCCATG CCTTGCAAAA TGGCGTTACT TAAGCTAGCT TGCCAAACCT 300
45 ACGGCTGGGG TCTTTCACGT TTATATCTAT GAGGGGAAGG ACCCAGAGTG GGAAGCTGG 360
GATCTTGGGA ACACGCTTCT CTACATGGCA TTGTCTGCAC GGTGGAGTCC GGATCTGAGC 420
TTGGCTTGGT TTTTAAAACC AGCCTGGAGT AGAGCAGATG GGTTAAGGTG AGTGACCCCT 480
50 CAGCCCTGGA CATTCTTAGA TGAGCCCCCT CAGGAGTAGA GAATAATGTT GAGATGAGTT 540
CTGTTGGCTA AAATAATCAA GGCTAGTCTT TATAAACTG TCTCCTCTTC TCCTAGCTTC 600
55 GATCCAGAGA GAGACCTGGG CGGAGCTGGT CGCTGCTCAG GAACTCCAGG AAAGGAGAAG 660

	CTGAGGTTAC CACGCTGCGA ATGGGTTTAC GGAGATAGCT GGCTTTCCGG GGTGAGTTCT	720
	CGTAAACTCC AGAGCAGCGA TAGGCCGTAA TATCGGGGAA AGCACTATAG GGACATGATG	780
5	TTCCACACGT CACATGGGTC GTCCTATCCG AGCCAGTCGT GCCAAAGGGG CGGTCCCGCT	840
	GTGCACACTG GCGCTCCAGG GAGCTCTGCA CTCCGCCCCG AAAGTGCGCT CGGCTCTGCC	900
10	AGGACGCGGG GCGCGTGA CTATGCGTGGGC TGGAGCAACC GCCTGCTGGG TGCAAACCCT	960
	TTGCGCCCCG ACTCGTCCAA CGACTATAAA GAGGGCAGGC TGTCTCTTAA GCGTCACCAC	1020
	GACTTCAACG TCCTGAGTAC CTTCTCCTCA CTTACTCCGT AGCTCCAGCT TCACCAGATC	1080
15	CTCGAGAACG TCTCCCATGG GCTATCCCTA TGACGTCCCG GATTACGCAG TCATGGGCAG	1140
	CAGCCATCAT CATCATCATC ACAGCAGCGG CCTGGTGCCG CGCGGCAGCC ATATGGATCA	1200
20	GAACAACAGC CTGCCACCTT ACGCTCAGGG CTTGGCCTCC CCTCAGGGTG CCATGACTCC	1260
	CGGAATCCCT ATCTTTTAGTC CAATGATGCC TTATGGCACT GGACTGACCC CACAGCCTAT	1320
	TCAGAACACC AATAGTCTGT CTATTTTGGA AGAGCAACAA AGGCAGCAGC AGCAACAACA	1380
25	ACAGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGCAACAG CAGCAGCAGC AGCAGCAGCA	1440
	GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA GCAACAGGCA GTGGCAGCTG CAGCCGTTCA	1500
30	GCAGTCAACG TCCCAGCAGG CAACACAGGG AACCTCAGGC CAGGCACCAC AGCTCTTCCA	1560
	CTCACAGACT CTCACAACTG CACCCTTGCC GGGCACCCT CCACTGTATC CCTCCCCCAT	1620
	GACTCCCATG ACCCCCATCA CTCCTGCCAC GCCAGCTTCG GAGAGTTCTG GGATTGTACC	1680
35	GCAGCTGCAA AATATTGTAT CCACAGTGAA TCTTGTTGTG AAATTGACC TAAAGACCAT	1740
	TGCACTTCGT GCCCGAAACG CCGAATATAA TCCCAAGCGG TTTGCTGCGG TAATCATGAG	1800
40	GATAAGAGAG CCACGAACCA CGGCACTGAT TTTCAGTTCT GGGAAAATGG TGTGCACAGG	1860
	AGCCAAGAGT GAAGAACAGT CCAGACTGGC AGCAAGAAAA TATGCTAGAG TTGTACAGAA	1920
	GTTGGGTTTT CCAGCTAAGT TCTTGGACTT CAAGATTGAG AACATGGTGG GGAGCTGTGA	1980
45	TGTGAAGTTT CCTATAAGGT TAGAAGGCCT TGTGCTCACC CACCAACAAT TTAGTAGTTA	2040
	TGAGCCAGAG TTATTTCTCTG GTTTAATCTA CAGAATGATC AAACCCAGAA TTGTTCTCCT	2100
50	TATTTTGTGT TCTGGAAAAG TTGTATTAAAC AGGTGCTAAA GTCAGAGCAG AAATTTATGA	2160
	AGCATTTGAA AACATCTACC CTATTCTAAA GGGATTGAGG AAGACGACGT AATGGCTCTC	2220
	ATGTACCCTT GCCTCCCCCA CCCCCTTCTT TTTTTTTTTT TAAACAAATC AGTTTGTGTTT	2280
55	GGTACCTTTA AATGGTGGTG TTGTGAGAAG ATGGATGTTG AGTTGCAGGG TGTGGCACCA	2340

GGTGATGCCC TTCTGTAAGT GCCCCTTCCG GCATCCCGGA ATTCCTGCAG CCCAACGCGG 2400
 CCGCTTCGAG GGATCTTTGT GAAGGAACCT TACTTCTGTG GTGTGACATA ATTGGACAAA 2460
 5 CTACCTACAG AGATTTTAAAG CTCTAAGGTA AATATAAAAT TTTTAAGTGT ATAATGTGTT 2520
 AAACACTGA TTCTAATTGT TTGTGTATTT TAGAFTCCAA CCTATGGAAC TGATGAATGG 2580
 10 GAGCAGTGGT GGAATGCCTT TAATGAGGAA AACCTGTTTT GCTCAGAAGA AATGCCATCT 2640
 AGTGATGATG AGGCTACTGC TGA CTCTCAA CATTCTACTC CTCCAAAAAA GAAGAGAAAG 2700
 GTAGAAGACC CCAAGGACTT TCCTTCAGAA TTGCTAAGTT TTTTGAGTCA TGCTGTGTTT 2760
 15 AGTAATAGAA CTCTTGCTTG CTTTGCTATT TACACCACAA AGGAAAAAGC TGCACTGCTA 2820
 TACAAGAAAA TTATGGAAAA ATATTCTGTA ACCTTTATAA GTAGGCATAA CAGTTATAAT 2880
 CATAACATAC TGTTTTTTCT TACTCCACAC AGGCATAGAG TGTCTGCTAT TAATAACTAT 2940
 20 GCTCAAAAAT TGTGTACCTT TAGCTTTTTA ATTTGTAAAG GGGTTAATAA GGAATATTTG 3000
 ATGTATAGTG CCTTGACTAG AGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT 3060
 25 TGCTTTAAAA AACCTCCAC ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT 3120
 TGTTGTTAAC TTGTTTATTG CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA 3180
 30 TTTCACAAAT AAAGCATTTT TTTCACTGCA TTCTAGTTGT GGTGTGTCCA AACTCATCAA 3240
 TGTATCTTAT CATGTCTGGA TCC 3263

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 371 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

- (ix) FEATURE:
 (A) NAME/KEY: Protein
 (B) LOCATION: 1..371

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Val Met Gly Ser Ser
 1 5 10 15
 His His His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser His
 20 25 30

325 330 335

Phe Val Ser Gly Lys Val Val Leu Thr Gly Ala Lys Val Arg Ala Glu
 340 345 350

5 Ile Tyr Glu Ala Phe Glu Asn Ile Tyr Pro Ile Leu Lys Gly Phe Arg
 355 360 365

10 Lys Thr Thr
 370

(2) INFORMATION FOR SEQ ID NO: 17:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

25 (ix) FEATURE:
 (A) NAME/KEY: Protein
 (B) LOCATION: 1..18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

30 Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
 1 5 10 15

Arg Gly

026083-0173